Predictive Analytics and the Targeting of Audits*

Nigar Hashimzade
Durham University and Tax Administration Research Centre

Gareth D. Myles†
University of Exeter and Tax Administration Research Centre

Matthew D. Rablen
Brunel University London and Tax Administration Research Centre

October 22, 2015

Abstract

The literature on audit strategies has focused on random audits or on audits conditioned only on income declaration. In contrast, tax authorities employ the tools of predictive analytics to identify taxpayers for audit, with a range of variables used for conditioning. The paper explores the compliance and revenue consequences of the use of predictive analytics in an agent-based model that draws upon a behavioral approach to tax compliance. The taxpayers in the model form subjective beliefs about the probability of audit from social interaction, and are guided by a social custom that is developed from meeting other taxpayers. The belief and social custom feed into the occupational choice between employment and two forms of self-employment. It is shown that the use of predictive analytics yields a significant increase in revenue over a random audit strategy.

Keywords: tax compliance; social network; agent-based model.
JEL: H26; D85.

---

*We thank two guest editors; an anonymous referee; the participants of the Taxation, Social Norms and Compliance Conference (Nuremberg) and the Shadow Economy, Tax Evasion and Governance Conference (Münster); and seminar audiences at Brunel and Tulane for their helpful comments. We are grateful to the ESRC for financial support under grant ES/K005944/1.

†Department of Economics, University of Exeter, Exeter EX4 4PU, UK. e-mail: gdmyles@ex.ac.uk

1 Introduction

The standard analysis of tax compliance in Allingham and Sandmo (1972) and Yitzhaki (1974), and much of the literature that has followed, is based on the assumption that taxpayers abide by the axioms of expected utility theory and that audits are random. An exception is the literature on optimal auditing – including Reinganum and Wilde (1985, 1986) and Chander and Wilde (1998) – which characterizes the equilibrium audit strategy as a function of reported income. In practice, the overwhelming majority of audits performed by tax authorities are "risk-based" (in which taxpayers are targeted for audit), with only a small fraction of audits performed on a random basis for statistical purposes. Unlike the presumption of the optimal auditing literature, however, the targeting of risk-based audits is not based solely on the income report. Rather, tax authorities rely on the experience of case officers reviewing returns and, increasingly, on the basis of predictive analytics which applies statistical tools to the data on a set of taxpayers' characteristics, often in the form of qualitative variables (see Cleary, 2011; Hsu et al., 2015). The expected utility model has also been subject to significant criticism and many alternatives models with behavioral foundations have been proposed.

The paper explores the compliance and revenue consequences of the use of predictive analytics in an agent-based model that draws upon a behavioral approach to tax compliance. We use agent-based modelling because this allows us to explore a richer model than is possible in a theoretical analysis but means we rely on simulation to generate our results. The model is constructed on the foundation of a social network that governs the interaction between taxpayers and the transmission of information between taxpayers. The information consists of attitudes towards compliance (in the form of a social custom) and beliefs about audits (a subjective probability of being audited). Taxpayers must make an occupational choice between employment and two forms of self-employment. Employment provides a safe income but because of the third-party reporting of income there is no possibility of non-compliance. The two self-employment occupations are risky, but non-compliance is possible. Risk averse taxpayers allocate into the occupations so as to maximize expected utility given the expected income and riskiness of each occupation. Given the different levels of risk in the occupations, taxpayers divide among occupations on the basis of risk aversion. This results in self-selection of those who will exploit opportunities for non-compliance into occupations where such opportunities arise.

The predictive analytics investigated in the model are based on Tobit and logit regression models using data from tax returns and from the outcomes of past audits. The Tobit model targets audits on the basis of predicted evasion level and the logit model on the basis of predicted likelihood of non-compliance. The predictive analytics are implemented by simulating the model with random audits for an initial period to acquire audit data, and then using this data to target audits where non-compliance is predicted. It is shown that predictive analytics secure a significant increase in revenue over a random audit strategy.

To give the results validity it is necessary to build the agent-based model on a sound underlying theory of the compliance decision. Our modelling starts from the assumption that taxpayers do not know the audit strategy of the tax authority but must form a belief about the probability of being audited. This is consistent with the idea of behavioral economics that individuals generally do not evaluate risky prospects using the objective probabilities of events but form subjective probabilities (or transform objective probabilities using a weighting function). The subjective probabilities (or, in our terminology, *beliefs*) can differ significantly, and persistently, from the objective probability (Kahneman and Tversky, 1979). There is also empirical (Spicer and Lundstedt, 1976) and experimental (Baldry, 1986) evidence that the individual compliance decision also takes into account social factors such as the perceived extent of evasion in the population. We choose to summarize the range of social factors as the *attitude* of the taxpayer toward compliance. This is essentially identical to the concept of *tax morale* that is prominent in the empirical literature (e.g., Torgler, 2002).

A key feature of our modelling is to make explicit the processes through which the attitude towards compliance and the belief about auditing are formed. Attitudes and beliefs are endogenous and result from the interaction of a taxpayer with other taxpayers and with the tax authority. The importance of interaction makes it necessary to specify the social environment in which the interaction takes place. We do this by employing a social network with a given set of links between taxpayers to govern the flow of information. After each round of audits some of the taxpayers who are linked will meet and exchange information. The likelihood of information transmission is greater between taxpayers in the same occupation.

The paper is structured as follows. Section 2 describes the separate concepts that are built into the model. Section 3 provides analytical details on how these concepts are implemented. Sections 4 and 5 describe the simulation results under a random audit rule and when the audit rule is informed by predictive analytics. Section 6 concludes.

2 Conceptual Approach

This section describes the elements that constitute the agent-based model. The purpose of the discussion is to relate these elements to the extensive literature on the individual tax compliance decision. The seminal analyses of the compliance decision by Allingham and Sandmo (1972) and Yitzhaki (1974) were built upon the application of expected utility theory. A standard criticism of this model is that it over-predicts the extent of evasion when evaluated using the objective probability of audit which has motivated the application of ideas from behavioral economics.[1] The behavioral models of the compliance decision are surveyed in Hashimzade et al. (2013).

The key elements of our agent-based model is that taxpayers make an *occupational choice* decision prior to the compliance decision. The compliance decision is based on the attitude toward compliance as summarized in a *social custom* and belief about audits captured in a *subjective probability* of being audited. The information used to form attitudes and beliefs

[1] It should be noted that Slemrod (2007) gives good reasons why this claim should be treated with caution.

is transmitted through meetings between taxpayers governed by a *social network*. Each of these components is now described in greater detail.

2.1 Occupational choice

Occupational choice determines the possibility for engaging in tax evasion. Income from employment is often subject to a withholding tax and/or third-party reporting to the tax authority. For example, the UK Pay-As-You-Earn system involves income tax being deducted by employers and remitted directly to the tax authority. This prevents evasion by employees (unless there is collusion with the employer) and so non-compliance is only possible for taxpayers who are self-employed. Occupations also differ in their traditions concerning payment in cash. Those in which cash payment is common provide the greater opportunity for evasion. Occupational choice has not had a prominent role in the literature on tax evasion despite its clear importance. Exceptions to this are Pestieau and Possen (1981) who model occupational choice, Cowell (1981), Isachsen and Strøm (1980), and Trandel and Snow (1999) who analyze the choice between working in the regular and the informal economy.

Occupational choice is also important for the connection it has with risk aversion. Individuals allocate to occupations on the basis of their ability at that occupation and their attitude to risk. Those who are least risk averse will choose to enter the riskiest occupations. Kanbur (1979) and Black and de Meza (1997) assume employment is safe but self-employment is risky, and address the social efficiency of aggregate risk-taking. Self-employment attracts the least risk-averse taxpayers, who will evade the most when the opportunity arises. Hence, occupational choice has the effect of self-selecting taxpayers who will evade into a situation in which they can evade. This observation should form part of any explanation of why non-compliance can be so significant within specific occupational groups.

Our model includes a choice between employment and two forms of self-employment. Employment is a safe activity that delivers a certain income. Self-employment is risky so that each taxpayer only knows the probability distribution of income when making an occupational choice. One of the self-employment occupations is riskier than the other, in a sense we make precise below.[2]

2.2 Social customs

The experiments of Baldry (1986) provide compelling evidence that the evasion decision is not just a simple gamble. This can be rationalized by introducing an additional cost into the evasion decision. These costs can be financial (Bayer, 2006; Chetty, 2009; Lee, 2001) or psychic (Gordon, 1989). Psychic costs can arise from fear of detection or concern about the shame of being exposed. The magnitude of the psychic cost can reflect an individual's attitude towards compliance. Attitudes are an important feature of psychological theories of tax compliance (Kirchler et al., 2008; Weigel et al., 1987). The psychic cost can also be

[2] Individuals differ in their level of skill in the occupations, and skills are one of the determinants of income. This makes it necessary to state the formal details before "riskier" can be explained in full.

interpreted as the loss of the payoff from following a social norm for honest tax payment. Adopting this interpretation makes it natural to assume that the size of the loss in payoff is generated by explicit social interaction, and that the size is larger when fewer taxpayers evade (Fortin et al., 2007; Kim, 2003; Myles and Naylor, 1996; Traxler, 2010).

The additional costs have an important role in explaining some features of the tax evasion decision. We model attitudes by including a social custom of honest tax payment in the model so that there is a utility gain (relative to the state with non-compliance) when tax is paid in full. The importance attached to the social custom by each taxpayer is determined by their interaction with other taxpayers within the social network.

2.3 Subjective beliefs

We have already observed that if choice is based on objective probability of being audited then the standard model over-predicts the amount of evasion. This has led to the application of choice models based on non-expected utility theories. Non-expected utility models can predict the correct level of evasion for reasonable parameter values. This is because they permit the subjective probability of audit (the weighting on the payoff when audited) to be greater than the objective probability. They also open the possibility of designing compliance policy to manipulate the subjective nature of the decision (Elffers and Hessing, 1997).

It is standard to distinguish between a choice with risk (the agent knows the probability distribution of future events) and a choice with uncertainty (the agent does not know the probabilities). A first step away from expected utility theory is to consider a choice with risk but to assume the probabilities are distorted into "decisions weights" that enter the expected payoff. Rank-dependent expected utility (Quiggin, 1981, 1982; Quiggin and Wakker, 1994) uses a particular weighting scheme to transform the objective probability of events into subjective probabilities and has been applied to the evasion decision by Arcand and Graziosi (2005), Bernasconi (1998) and Eide (2001). Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) also uses a weighting scheme but payoffs are determined by gains and losses relative to a reference point. It has been applied to compliance by al-Nowaihi and Dhami (2007), Bernasconi and Zanardi (2004), Rablen (2010), and Yaniv (1999). Uncertainty has been modelled by assuming the agent forms a probability distribution over the possible probability distributions of outcomes ("second-order uncertainty"). This gives rise to the concept of ambiguity, surveyed in Camerer and Weber (1992) which has been applied to tax compliance by Snow and Warren (2005).

We incorporate these ideas into the model by assuming that each taxpayer forms a subjective belief about the audit probability and by explicitly modelling the process for forming beliefs. This allows the model to provide an explanation of how subjective probabilities can endogenously emerge and remain systematically different from the objective probabilities.

2.4 Social network

The illegality of tax evasion and the incentive the tax authority has to conceal its audit strategy imply that taxpayers cannot be fully informed. A natural assumption is that information will not be revealed publicly, but will be transmitted between taxpayers in a position of mutual trust. The social network we adopt is a formalization of this assumption.

The importance of social contacts is supported by empirical evidence on the positive connection between the number of tax evaders a taxpayer knows and the extent of evasion of that taxpayer (De Juan et al., 1994; Geeroms and Wilmots, 1985; Spicer and Lundstedt, 1976; Wallschutzky, 1984; Webley et al., 1988). This evidence demonstrates that the compliance decision is not made in isolation but that each taxpayer makes reference to the observed behavior of the society in which they operate.

We capture this social interaction by applying network theory (Goyal, 2009; Jackson, 2004). Networks have previously been used in the analysis of evasion by Korobow et al. (2007) and Franklin (2009). They have also been applied to crime more generally (Glaeser et al., 1996).

The social network in our model plays two roles. First, it transmits the social custom from one person to another: if two non-evaders meet the importance of the social custom of honest payment is increased for both, but if a non-evader meets an evader then it is reduced for the non-evader and increased for the evader. Second, the network transmits information about audit policy. Since the audit strategy is not public information, taxpayers have to infer it from their own experience and from the receipt of information about the experiences of others. Our simulations are an application of agent-based modelling (Bloomquist, 2004; Tesfatsion, 2006) with agent interaction controlled by the social network.

3 Network Model

In this section we model the formation of attitudes and beliefs as the outcome of social interaction, and opportunities as the outcome of occupational choice. This is achieved by applying the theory of network formation to track the links between taxpayers and the transmission of attitudes and beliefs, and combining this with agent-based modelling which employs a behavioral approach to describe individual choices.

There are N individuals, indexed by $j = 1, ..., N$, whose lives extend throughout the simulation. Although the lives are long, each individual makes a succession of single-period decisions and so is "myopic". We discuss possible relaxation of this assumption in Section 6. Individuals interact repeatedly in discrete time, $t = 0, ..., T$, where each period t is understood to be the tax return period. Each individual, j, at time t is described by a vector of characteristics

$$\left(w_j, \rho_j, s_j^1, s_j^2, z_j; p_{j,t}^1, p_{j,t}^2, \chi_{j,t}\right). \tag{1}$$

At the start of a simulation the values for all characteristics are randomly assigned to each taxpayer by making draws from independent distributions. Once drawn, the first five characteristics in (1) remain constant throughout the simulation and so represent the *exogenous* parameters describing the agents. These characteristics are w_j, the wage in employment (occupation 0); ρ_j, the coefficient of relative risk aversion; s_j^{α}, the skill in self-employed occupation α, $\alpha = 1, 2$; and z_j, the payoff from following the social custom. To avoid the potential for misleading results that might follow from an unusual draw of these characteristics our results are calculated as the average of multiple independent simulations. In particular, we simulate five times to obtain the baseline results in Section 4 and ten times for the analysis in Section 5.

The remaining three characteristics in (1) are *endogenous* and so are updated period-by-period through interaction with the tax authority and with other taxpayers in the social network. These characteristics are additionally indexed by time, t. These are: $p_{j,t}^{\alpha} \in [0, 1]$, the perceived (subjective) probability of audit in occupation α, $\alpha = 1, 2$, and $\chi_{j,t} \in [0, 1]$, the weight attached to the payoff from following the social custom for honesty.

We now describe how these characteristics enter into the choice problem of a taxpayer and how the subjective probability and weight given to social custom are updated.

In each period, t, every individual chooses their preferred occupation and, once income is realized, the optimal level of evasion. Individual j has a choice between employment or entering one of two self-employment occupations. These are occupations 1 and 2, which we denote by SE1 and SE2.[3] If employment is chosen the wage, w_j, is obtained with certainty, whereas in self-employment earnings are random. The outcome of self-employment for individual j in occupation α at time t is given by $s_j^{\alpha} y_{j,t}^{\alpha}$, where s_j^{α} is drawn once at the beginning of the simulation, but $y_{j,t}^{\alpha}$ is drawn randomly at every time t from the probability distribution function $F^{\alpha}(\cdot)$.[4] The choice of occupation is taken on the basis of $F^{\alpha}(\cdot)$ but the choice of evasion is made after the realization of $y_{j,t}^{\alpha}$. It is assumed that $E(y^1) < E(y^2)$ and $Var(y^1) < Var(y^2)$, so if $s_j^1 = s_j^2$ SE2 is riskier than SE1 but offers a higher expected income. Both self-employment occupations are riskier than employment, in the sense that for each agent the wage in employment is certain, i.e., $Var(w_j) = 0$.

It is not possible to evade tax in employment because income is subject to third-party reporting or to a withholding tax. Evasion only becomes possible when self-employment is chosen. In occupation α taxpayer j has belief at time t that the probability of being audited and evasion being detected[5] is $p_{j,t}^{\alpha}$. The belief about the audit probability is updated through the experience of the taxpayer with audits and through the exchange of information when meeting other taxpayers. The attitude of taxpayer j toward evasion is summarized in $\chi_{j,t}$, the weight given to the social custom. This attitude is also updated through meetings with other taxpayers. We describe the processes for updating attitudes and beliefs in detail after

[3] It may seem unrealistic to have an occupational choice in every period but in the simulations only a very small proportion of taxpayers actually change occupation in any period.

[4] For tractability, we abstract in this setting from the possibility of bankruptcy for agents in the self-employment occupations. Were this feature allowed for (see, for example, Bloomquist, 2011), business survival would become a function of the outcome of the tax evasion gamble, which generates a dynamic effect on compliance in addition to those we study in our model. We are grateful to an anonymous referee for this point.

[5] Here we assume that detection is full. Alternatively, one can assume partial detection, which gives rise to a number of interesting issues which are beyond the scope of this paper.

discussing the choice of occupation for given attitudes and beliefs.

The choice of occupation and the choice to evade tax involve risk. Taxpayer j has a (constant) degree of relative risk aversion measured by the risk aversion parameter, ρ_j. The taxpayer chooses occupation and evasion level at time t to maximize subjective expected utility given beliefs $\{p_{j,t}^{\alpha}\}$. For analytical tractability, we assume throughout a CRRA form for utility:

$$U_j(Y) = \frac{Y^{1-\rho_j} - 1}{1-\rho_j}. \tag{2}$$

The attitude toward evasion determines the utility value of following the social custom of honest tax payment. The payoff from the social custom is given by z_j and the individual weight, or the importance, assigned to this payoff by the taxpayer is determined by $\chi_{j,t}$. Hence, compliance with tax payment at time t generates an additional utility from following the social custom of $\chi_{j,t} z_j$.

In employment there is no opportunity for evasion, and so the taxpayer obtains a payoff given by

$$V^0 = \frac{[(1-\tau)w_j]^{1-\rho_j} - 1}{1-\rho_j} + \chi_{j,t} z_j,$$

where τ is the constant marginal tax rate. The possibility of tax evasion makes the choice of self-employment a compound lottery: the income is random, as is the outcome of choosing to evade.

Define the expected payoff from the optimal choice of evasion in self-employment occupation α for a given realization $y_{j,t}^{\alpha}$ as

$$\begin{aligned} V_e^{\alpha}\left(y_{j,t}^{\alpha}\right) \ = \ & \max_{E \in \left[0, s_j^{\alpha} y_{j,t}^{\alpha}\right]} \left\{ p_{j,t}^{\alpha} U_j\left(\left[1-\tau\right] s_j^{\alpha} y_{j,t}^{\alpha} - f\tau E\right)\right. \\ & \left. + (1-p_{j,t}^{\alpha}) U_j\left(\left[1-\tau\right] s_j^{\alpha} y_{j,t}^{\alpha} + \tau E\right) + \chi_{j,t} z_j \mathbf{1}_{[E=0]} \right\}, \end{aligned}$$

where $f > 1$ is the fine levied on unpaid tax if evasion is detected. The term $\mathbf{1}_{[A]}$ is an indicator function that takes the value of one if A is true and zero otherwise: the payoff from the social custom is obtained only if tax is paid in full. The level of evasion, $E_{j,t}^{\alpha} = E\left(y_{j,t}^{\alpha}\right)$, will be a function of the realized income $y_{j,t}^{\alpha}$ in occupation α. The expected payoff from the compound lottery describing occupation α is then

$$V^{\alpha} = \int V_e^{\alpha}(y)\, dF^{\alpha}(y).$$

The choice of occupation is made by comparing the (expected) utility levels from employment and from self-employment. Hence, the chosen occupation is given by selecting the maximum of $\{V^0, V^1, V^2\}$.

After self-employment occupation α is chosen at time t an outcome $\tilde{y}_{j,t}^{\alpha}$ is realized according to the probability distribution function $F^{\alpha}(\cdot)$. Given the outcome, the optimal evasion decision is implemented, as described above. Denote the level of evasion that is realized by $\tilde{E}_{j,t}^{\alpha} = E\left(\tilde{y}_{j,t}^{\alpha}\right)$. Tax returns are submitted, and a proportion of those in self-employment occupations are then audited, according to a rule chosen by the tax authority. If evasion is discovered, unpaid tax is reclaimed and the fines on unpaid tax are paid.

The social network is modelled as a set of bidirectional links described by an $N \times N$ symmetric matrix of zeros and ones. For example, in the network described by matrix A

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

the first row, representing the links of individual 1, has a single 1 in column 2 which means that 1 is linked to 2. There is a corresponding 1 in the first column in the second row representing the link of individual 2 with 1. That is, the element in row i and column j of matrix A is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are linked in the network,} \\ 0 & \text{otherwise.} \end{cases}$$

We use a random (Erdös–Rényi) network, where agents i and j are linked with a common exogenous probability, $\Pr[A_{ij} = 1] = \nu$; the matrix is created at the outset and does not change.[6] The network determines who may meet whom to exchange information. In each period a random selection of meetings occur described by a matrix C^t of zeros and ones randomly drawn in every period. Individuals i and j meet during period t if $A_{ij}C^t_{ij} = 1$; at a meeting they may or may not exchange information about their subjective probability of audit in each self-employed occupation and about whether each of them was compliant in that period. The probability of information exchange depends on the occupational groups to which i and j belong; the probability is highest when they are in the same occupation.

Recall that individuals in all three occupations hold beliefs about the probability of being audited in each of the two self-employment occupations; they also know that evasion is not possible in employment. We assume there are two ways in which beliefs are updated. Consider taxpayer j who has worked in occupation $\alpha \in \{0, 1, 2\}$ in period t. If j is self-employed, after submission of the tax return this taxpayer may or may not be audited. On the basis of the outcome j's belief about the audit probability, $p^{\alpha}_{j,t}$, in that occupation is then adjusted. The beliefs about the audit probability in the other self-employment occupation, $p^{\beta}_{j,t}$, $\beta \neq \alpha$, remain unchanged at this stage. Following this, the taxpayer may meet with a contact in the network. Let the meeting be with a taxpayer i who is engaged in occupation $\beta \in \{0, 1, 2\}$. At the meeting information is exchanged with probability $q^{\alpha\beta}$. This information is then used to update the belief about the audit probabilities, $p^{\gamma}_{j(i),t}$, $\gamma = \{1, 2\}$, in both self-employment occupations. If taxpayer j is in paid employment then his or her

[6] Here the network is fixed, but the probabilities of information exchange between the linked individuals change if they switch occupations, as we shall describe. Another possibility would be to have the network itself revised as a consequence of chosen actions, i.e. agents in different occupations belonging to different social networks. The model may also be analysed under, for instance, the small-world or power law networks as alternatives to the random network (see, e.g., Andrei et al., 2014; Axtell et al., 2006; Souma et al., 2003.)

belief about audit probabilities in self-employment, $p^{\gamma}_{j,t}$ $\gamma = \{1, 2\}$, can only be updated at the meeting.

The choice of occupation in period $t + 1$ is made on the basis of the beliefs $\left\{p^1_{j,t}, p^2_{j,t}\right\}$ updated after the audits and the information exchange. Two different processes for the updating of subjective beliefs following an audit have been proposed in the literature. As studies have reliably demonstrated important deviations from Bayesian inference (see, for example, Grether, 1980), we allow for non-Bayesian updating. The first process, which is qualitatively similar to a Bayesian process, is to assume that individuals feel marked as targets if they are audited, so that one audit is believed likely to be followed by another. We term this the *target effect.* In contrast, those not audited in one period believe they are less likely to be audited in the next period. Formally, if taxpayer j in occupation $\alpha \in \{1, 2\}$ is audited in period t, his or her belief about being audited in the same occupation in the next period is raised to probability P, otherwise it decays. The updating rule for the subjective probability is therefore

$$\tilde{p}^{\alpha}_{j,t+1} = \begin{cases} P \in [0, 1] & \text{if audited at } t, \\ \delta p^{\alpha}_{j,t+1},\ d \in [0, 1] & \text{otherwise.} \end{cases} \tag{3}$$

We refer to the case of $P = 1$ as the maximal target effect. The alternative is the *bomb-crater* effect that is documented experimentally by Guala and Mittone (2005), Kastlunger et al. (2009), Maciejovsky et al. (2007), and Mittone (2006). In this process a taxpayer who has been audited in one period believes that they are less likely to be audited in the next, but the belief gradually rises over time. The process is therefore described by

$$\tilde{p}^{\alpha}_{j,t+1} = \begin{cases} P \in [0, 1] & \text{if audited at } t, \\ p^{\alpha}_{j,t} + \delta\left(1 - p^{\alpha}_{j,t+1}\right),\ \delta \in [0, 1] & \text{otherwise,} \end{cases} \tag{4}$$

with $P = 0$ being the maximal bomb-crater effect.

Recent empirical analysis using administrative data for the entire population of UK taxpayers who submit tax returns has produced results that are in agreement with the target effect. Advani et al. (2015) report that compliance increases post-audit and remains high for several periods before beginning to decrease again. This matches the target effect but is the opposite of what the bomb-crater effect predicts. Since the empirical results are based on analysis of actual behavior of the entire population of self-reporting taxpayers they are very convincing, and so we adopt the target effect in the simulations that follow.[7]

After the audit process is completed a taxpayer may meet with a contact in their social network. The information that may (or may not) be exchanged at a meeting includes the subjective probabilities and whether or not the agents were audited. If taxpayer j meets taxpayer i and if the information exchange takes place, j's subjective probability of audit in occupation α is updated according to the rule

$$p^{\alpha}_{j,t+1} = \mu\tilde{p}^{\alpha}_{j,t} + (1 - \mu)\,\tilde{p}^{\alpha}_{i,t},\ \alpha = \{1, 2\}\,.$$

[7] We considered the bomb-crater effect in earlier versions of the paper (see Hashimzade et al., 2014). The main results on the effects of predictive analytics are qualitatively similar to those we present here under the target effect.

It is reasonable to assume that the probabilities of information exchange depend on whether or not the taxpayers belong to the same occupation. In the numerical simulations we consider two cases, labelled "the focussed information transmission" and "the diffused information transmission", described in detail in Section 4.

The importance assigned to the social custom is also determined by interaction in the social network. The weight, $\chi_{j,t}$, is updated in period t if information exchange occurs between j and some other taxpayer in that period. Assume individual j meets individual i in occupation α at time t and information exchange takes place. The updating process is described by a function, $\chi_{j,t+1} = g\left(\chi_{j,t}, \mathbf{1}_{\left[\tilde{E}^{\alpha}_{i,t}=0\right]}\right)$, such that (i) $\chi_{j,t+1} \geq \chi_{j,t}$ if information is exchanged with a compliant taxpayer, i.e., if $\mathbf{1}_{\left[\tilde{E}^{\alpha}_{i,t}=0\right]} = 1$; and (ii) $\chi_{j,t+1} < \chi_{j,t}$ if information is exchanged with an evader, i.e., if $\mathbf{1}_{\left[\tilde{E}^{\alpha}_{i,t}=0\right]} = 0$. In the simulations we assume a partial adjustment process, where j's individual weight adjusts by fraction $\varphi \in (0,1)$ to its lower bound (zero) if i was an evader and to its upper bound (one) if i was compliant:

$$\begin{aligned}\chi_{j,t+1} &= (1-\varphi)\,\chi_{j,t} + \varphi \mathbf{1}_{\left[\tilde{E}^{\alpha}_{i,t}=0\right]} \\ &= \begin{cases} (1-\varphi)\,\chi_{j,t} & \text{if } \tilde{E}^{\alpha}_{i,t} > 0, \\ \chi_{j,t} + \varphi\left(1-\chi_{j,t}\right) & \text{if } \tilde{E}^{\alpha}_{i,t} = 0. \end{cases}\end{aligned}$$

We assume that earnings in occupation α, $\alpha \in \{0,1,2\}$, are drawn from the lognormal distribution, $\log\mathcal{N}(\mu_\alpha, \sigma^2_\alpha)$, and that skills in self-employment are given by $\frac{1}{1-\gamma\varsigma}$, where ς is drawn from the standard uniform distribution and $\gamma \in (0,1)$ is a constant parameter. Each individual knows their own wage in employment, $y^0_{j,t} = w_j$, their own skill, $s^{1,2}_j$, in the self-employment occupations $\alpha = 1,2$, and the distribution of earnings, $F^\alpha(\cdot)$, in the self-employed occupations. At time $t = 0$ each individual is randomly assigned a vector of subjective beliefs, $\left\{p^1_{j,0}, p^2_{j,0}\right\}$, and the level of importance of social custom, $\chi_{j,0}$, independently from the standard uniform distribution. The objective probability of a random audit for all self-employed is 0.05; the employed are not audited.[8]

A key feature of the simulation model is the choice of occupation by taxpayers. We calibrate the model so that the allocation of taxpayers across occupations with random audits matches the allocation in the UK. The UK Living Costs and Food Survey (LCFS) is used to obtain data on the occupations of all responding households in 2012. Since this is a subsample of the UK population we use the annual weights provided by the LCFS to correct for sampling bias. The households reporting as self-employed are then separated into two occupations using the Alm and Erard (2013) classification of occupations into Non-Risky (our SE1) and Risky (our SE2) where there is known to be a higher presence of informal suppliers. This process provides baseline figures of 86 per cent in employment, eight per cent in SE1, and six per cent in SE2.

[8] According to the IRS, around 0.96 per cent of individual tax returns filed in calendar year 2012 were examined (IRS, 2014). The audit rate by the size of (adjusted gross) income varied from as low as 0.58 per cent for incomes between $75,000 and $100,000, to as high as 24.16 per cent for incomes above $10,000,000.

4 Baseline Simulations

We first conduct simulations of the network model described above under the assumption of random audits in order to obtain a baseline outcome. This allows an investigation of the nature of the equilibrium, a comparison of the model outcome with evidence, and the consequences of the alternative assumptions on information transmission between occupations.

The results we present assume the target effect for audits as specified in equation (3) as it is supported by recent empirical work (Advani et al., 2015). The results for the bomb-crater model differ only in the pattern of compliance after audit, as described below; the outcomes for the revenues are not qualitatively different from those under the target effect. A complete set of results for the bomb-crater model are available from the authors upon request. The values of the exogenous parameters and the distribution functions for the random variables are given in the Appendix.[9]

Two types of model are simulated that differ in the probability of information exchange between different groups. Let $q^{\alpha\beta}$ be the probability of information exchange between taxpayers in occupations α and β. The first model (denoted Foc) used focussed information transmission, $q^{\alpha\beta} = q^{\beta\alpha} = 0$ and $0 < q^{\alpha\alpha} \leq 1$ for $\alpha \neq \beta$, $\alpha, \beta = \{0, 1, 2\}$. That is, at a meeting information is exchanged with positive probability only between linked agents in the same occupation. The second model (denoted Diff) used diffused information transmission, $0 < q^{\alpha\beta} = q^{\beta\alpha} \leq q^{\alpha\alpha} \leq 1$ for $\alpha \neq \beta$, $\alpha, \beta = \{0, 1, 2\}$. In this case there is a positive probability that a meeting between linked taxpayers in different occupations results in information exchange and the probability of information exchange at meetings between members of the same occupation is reduced compared to that under the focussed information transmission. We perform ten independent simulations of each model over a 50-period horizon. Tables 1 and 2 report summary statistics under both focused and diffused information transmission. The results we report for each transmission mechanism are averages for the final period over the ten simulations performed. Figure 1 illustrates the results under diffused information transmission, calculated for each period as the average over the ten simulations performed.[10]

The central message from Figure 1 is that sub-groups of the population (the occupational groupings) can endogenously form different attitudes to compliance. As expected, the operation of self-selection sorts those who are most willing to accept risk into the riskiest occupation (SE2). Self-employment gives them the opportunity to evade, and they make use of this opportunity to become the least compliant group. The updating process for beliefs and the transmission of information around the social network result in the subjective probability of audit being above the true probability for the self-employed. The two self-employed

[9] The value of the social custom z is measured in units of utility. Therefore, although z appears constrained to take very small values, these values are commensurate with the values taken by the utility function in (2). Thus, with given parameterisation, a true report increases the utility of an "average" individual by about 10 per cent.

[10] The outcomes for each simulation were reasonably close; thus, across 50 periods the ratios of the standard deviation to the mean over ten simulations ranged from 0.8% to 5.8% for the risk aversion, 0.1% to 1.5% for social custom, 0.5% to 15.1% for beliefs, and 0.1% to 4.9% for compliance.

	Employment		SE1		SE2	
	Foc	Diff	Foc	Diff	Foc	Diff
Risk Aversion (ρ_j)	2.8540	2.8304	1.4070	1.4205	0.9762	0.9904
	(0.0140)	(0.0207)	(0.0629)	(0.0544)	(0.0549)	(0.0176)
Belief ($\chi_{j,t}$)	0.0002	0.0007	0.1731	0.2103	0.2057	0.2075
	(0.0001)	(0.004)	(0.0159)	(0.0191)	(0.0136)	(0.0098)
Compliance	1.0000	1.0000	0.7441	0.7199	0.6527	0.6305
	(0.0000)	(0.0000)	(0.0267)	(0.0202)	(0.0363)	(0.0258)

Table 1: The long-run characteristics by occupation under focussed (Foc) and diffused (Diff) information transmission (Diff). The numbers reported are the values for t=50 averaged across five independent simulations, with the sample standard deviations in parentheses.

groups hold similar beliefs, but these are distinctly different from those of the employed. The non-zero belief for the employed reflects their learning about audits from meeting with self-employed. The operation of the social custom results in the employed placing a high weight on following the custom for honesty. Taxpayers in the two self-employment occupations place much less weight on the social custom but this effect is not significantly different between the two occupations under the diffused information exchange. In contrast, with focused information exchange a significant difference in beliefs and weights placed on the social custom for honesty can emerge between the two self-employment occupations.

Table 1 reports the long-run values or risk aversion, beliefs and compliance for each occupational group (the sample means and the sample standard deviations for the final period, $t = 50$, from five independent simulations). The effect of self-selection into occupations is seen clearly in the mean level of risk aversion in each occupation. Risk aversion is significantly lower in SE2 than in SE1, and both are lower than in employment. The table also confirms that the subjective belief is above the true value of 0.05 of the probability of audit for those in self-employment, and that under focused information exchange it can differ between self-employment occupations. Compliance of those in employment is equal to one by construction, for this is measured as the proportion of agents who report truthfully. For both forms of information exchange compliance is lower for taxpayers in SE2.

Table 2 gives the population means and standard deviations (according to the assumed distribution) and the sample means and sample standard deviations for each occupational group for the wage in employment and skills in self-employment, as well as the sample means and sample standard deviations for the true earnings and reported earnings. The sample, again, consisted of the final periods ($t = 50$) of ten independent simulations. In interpreting the table it is important to note that the entries provide information on all three occupational possibilities a taxpayer could have chosen, not only the one they actually chose. For example, the table tells us that if, counterfactually, those taxpayers who actually chose SE2 had instead chosen SE1 their average skill in SE1 would have been 1.277. It can be seen that, once the individuals self-select into a particular occupation, the average productivity in each occupation (i.e., the wage for those in employment and skill for those

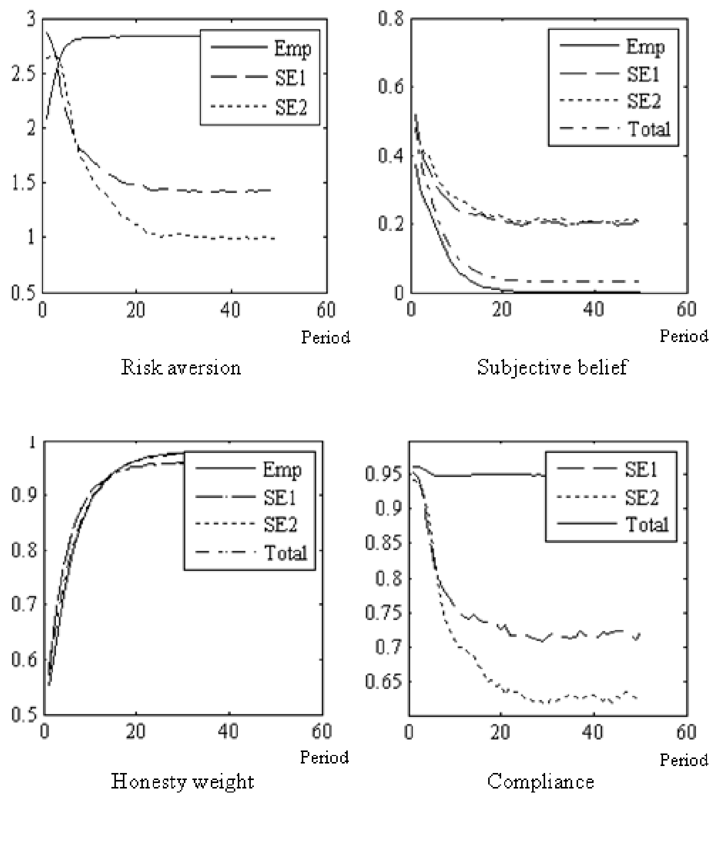

Figure 1: Risk aversion, subjective beliefs, social custom, and compliance of the employed (Emp), self-employed in occupations 1 (SE1) and 2 (SE2), and all taxpayers (All) under the diffused information transmission. The tax authority uses random audits.

Variable	Population	Employment		SE1		SE2	
		Foc	Diff	Foc	Diff	Foc	Diff
Wage (w_j)	13.045	13.3347	13.2723	11.3906	11.4092	11.6563	11.7109
	(2.00)	(0.0048)	(0.0054)	(0.0869)	(0.0777)	(0.0937)	(0.0959)
Skill in SE1 (s_j^1)	1.387	1.3518	1.3523	1.7533	1.7441	1.2966	1.277
	(0.28)	(0.0006)	(0.0007)	(0.0163)	(0.0168)	(0.0055)	(0.004)
Skill in SE2 (s_j^2)	1.387	1.3631	1.3677	1.3222	1.3385	1.7282	1.655
	(0.28)	(0.0004)	(0.0005)	(0.0032)	(0.0038)	(0.0238)	(0.004)
Declared earnings		13.3347	13.2723	9.5351	10.1166	8.8906	9.0089
		(0.0048)	(0.0054)	(0.2553)	(0.2837)	(0.3898)	(0.2962)
True earnings		13.3347	13.2723	14.0124	13.9719	14.1783	14.4470
		(0.0048)	(0.0054)	(0.2505)	(0.2874)	(0.4072)	(0.3579)

Table 2: Wages and skills in population and across occupations under focused (Foc) and diffused (Diff) information transmission. The second column describes the distributional assumptions. The remaining numbers are the values for t=50 averaged across ten independent simulations, with the sample standard deviations in parentheses.

in self-employment) is above the corresponding population mean, shown in the first column.

Hamilton (2000) estimates that, on average, the self-employed report income that is 35 per cent lower than in equivalent employment. In the simulation the reported incomes of the self-employed are between 25 per cent and 35 per cent lower than the employed, and so fit reasonably well with this observation. The evidence also suggests that reported income of the self-employed must be inflated by between around 29 per cent (Feldman and Slemrod, 2007), and 35 per cent (Pissarides and Weber, 1989) to account for under-reporting. It can be inflated to take account of the personal consumption of business goods by a further 34 per cent (Bradbury, 1997). The data from the simulation are, again, approximately consistent with these observations.

5 Random Audits and Predictive Analytics

The role of *predictive analytics* is to identify the best audit targets, in terms of the expected level or the expected likelihood of non-compliance. Predictive analytics are used by tax authorities in many countries, including the IRS and HMRC. The IRS, for instance, uses information from its random audit programme to design discriminant functions (DIF) that are used to assign a score – called the DIF score – to each tax return for the likelihood that it contains some irregularities or evasion. Various methods are used for identifying targets for risk-based audits. We wish to explore the effects of predictive analytics on compliance behavior and to examine the extent to which they can improve on other audit strategies.

The method we employ to conduct the analysis is to embed predictive analytics within the agent-based model. We then compare the outcome with predictive analytics based on tax return data to the outcome with random audits. The two forms of predictive analytics we investigate involve econometric analysis for predicting the level of non-compliance (level-targeting) and the probability of non-compliance (rate-targeting) by each taxpayer on the basis of the information provided on the tax return and by past audits.

This process is implemented in the simulation by using random audits (with each self-employed individual facing a 0.05 probability of being audited) for the first 50 periods to eliminate the effect of the initial conditions and to accumulate audit data. Data from the first ten periods are dropped from the statistics we record to remove starting point effects (but these ten periods are reported in the figures). The outcomes from the final five random audits are collected and at the end of period 50 are used to estimate a regression equation with the dependent variable being the amount of under-reported income in the first exercise and a binary variable taking the value of one if an individual under-reported their income and zero if reported truthfully in the second exercise.[11] The explanatory variables are the observed characteristics and the audit history of an agent; in our model these are termed *occupation*, *declaration*, and *previous audit*.[12] The estimated equation is used to predict non-compliance given information collected in period 51. From this point onward, the estimated regression equation in period t is used to predict non-compliance using tax return and audit history data in period $t+1$; the audit outcomes in period $t+1$ are added to the data set, and the regression analysis is repeated to predict non-compliance in $t+2$, and so on.

The simulations reported here compare a targeted regime in which audits are based entirely on predictive analytics with the outcome of random auditing. In each period that predictive analytics are used agents are sorted according to their predicted level of evasion (largest evaders in the first exercise) or according to their predicted probability of evasion (most likely evaders in the second exercise). In the targeted regime the top five per cent of taxpayers by predicted non-compliance are audited (this means that the number of targeted audits is equal to the average number of random audits, so that the audit costs are, on average, equal between these two strategies). In Hashimzade et al. (2014) we also report results on mixed regimes in which a combination of predictive analytics and random audits are used. The mixed regimes are used to explore the possibility that the agents learn about the audit patterns by exchanging information in their networks: if the audits concentrate on one occupation, individuals in the other occupation may evade more, having learned that they are likely to get away with it.

5.1 Targeting largest evaders

Since expected non-compliance is bounded below by zero (we do not allow for mistaken over-declaration) a Tobit (censored) regression is employed to predict the expected level of

[11] Alternatively, instead of under-reported income one can use unpaid tax as the variable of interest. In our model these approaches are equivalent because of the flat tax schedule.

[12] In the regression "previous audit" is a binary variable, one if the individual was audited in the previous period and zero otherwise. This specification can be easily extended to include a longer audit history and/or previous audit outcomes.

Variable	Coeff.	St. E.	$z-$stat.	ME (avg. data)	ME (indiv. avg.)
Const.	15.84	0.21	74.77	13.5338	13.9044
Declared Income	-4.89	0.67	-7.26	-4.1814	-4.2959
Previous audit	-35.73	144	-0.02	-30.5271	-31.3631
Self-employment 1	-1.28	1.49	-0.86	-1.0949	-1.1249

Table 3: Estimated coefficients and marginal effects in the evasion level equation (Tobit model). The reported coefficients and standard errors are the maximum likelihood estimates from the data accumulated up to the final period of one simulation (2077 observations).

evasion. We report the outcomes for the model with diffused information transmission; the results for the focused transmission are qualitatively the same. The parameter values are the same as in the baseline simulations (Section 4) and are listed in the Appendix.

Table 3 shows the marginal effects of explanatory variables upon the predicted level of under-reported income, calculated as the marginal effect for the average data point (*avg. data*) and as the average over individuals (*indiv. avg.*).The estimated coefficients change throughout each simulation as new data are added. To capture the long-run outcome we report the estimates based on the data accumulated up to the final period. To facilitate the reporting of standard errors, here we report the estimates from the last independent simulation performed, rather than an average across all the simulations performed.

One can see that individuals with higher declaration, those audited in the previous period, and those in SE1 are predicted to evade less tax (although only the declaration appears statistically significant). The latter implies that, *ceteris paribus*, the targeted audits will tend to focus on individuals in SE2. The lower evasion level of the previously audited taxpayers reflects the assumption of the target effect, as opposed to the bomb-crater effect (under the bomb-crater model the previous audit has a positive effect on evasion).

The simulation results when the largest evaders are targeted are shown in Figure 2. The values for each time period correspond to the averages over ten independent simulations.[13] The effect upon revenue of the introduction of predictive analytics is the sharp increase observed at period 50 in the first panel. This higher level of revenue is sustained for the remaining periods of the simulation. With random audits the subjective beliefs of the two self-employment occupations are the same. Once targeted audits are imposed the belief in SE2 is sustained just above the level under random auditing but falls in SE1. Compliance increases significantly in both self-employment occupations immediately after the imposition of predictive analytics, but then subsequently begins to fall in SE1 as taxpayers in this occupation learn that they are not the target of audits. This process continues until the point is reached at which the compliance of the two groups is approximately equal. This is close to optimal outcome: those in SE2 earn more on average than do those in SE1 so it is optimal that their compliance level is somewhat higher also.

[13] Across 90 periods, the ratios of the standard deviations to the means over ten simulations ranged from 0.2% to 0.9% for revenues, 0.1% to 4.9% for compliance, and 0.5% to 66.5% for beliefs.

The final panel of the figure shows the empirical cumulative distribution function (cdf) for revenue under random audits and under targeted audits. It can be seen that the cdf for targeted audits first-order stochastically dominates the cdf for random audits. This implies that expected revenue under targeting is higher. Moreover, for any objective function of the tax authority that is an increasing function of revenue the expected value of the objective function will be higher under targeting.

These results show clearly that the use of predictive analytics increases compliance and results in higher revenue. The increase in compliance raises the chance of a meeting with a compliant taxpayer and thus leads to a steady increase in the importance of social custom of honest reporting when predictive analytics are in operation. Compliance is not uniformly increased across occupational groups when predictive analytics are introduced because of the reduction in focus on the least compliant occupation, SE2. With targeted audits the proportion targeted towards taxpayers in SE2 is very high, as one can see in Figure 3; as a result, compliance among taxpayers in SE1 drops. Despite this, the results still demonstrate that a policy of targeted audits outperforms random auditing. Our findings do though point to the need to support predictive analytics with random audit programme to monitor the behavior of taxpayers outside the targeted group.[14]

5.2 Targeting most likely evaders

In the previous sub-section the predictive analytics targeted audits at those individuals with the highest predicted levels of non-compliance, or undeclared income. An alternative strategy is to target those predicted most likely to evade, or those with the highest evasion score. The evasion score is similar to the credit score used by credit-rating agencies, and can be calculated by estimating a logit or probit regression. For the explanatory variables we use the same observed characteristics of the agents and their audit histories as in the previous exercise.

In the data the evasion score is assigned the value of one if the individual evaded tax and zero if declared truthfully. The predicted values of the evasion score from the regression have the interpretation of the predicted probabilities that an individual will under-report their income. In every period the individuals are ranked according to their predicted evasion score, and those with the highest predicted score are audited. Again, the number of audits is equal to the average of the number of random audits, in order to equalize the audit costs across strategies. The targeted audits are then compared with random audits. We present the results obtained from the logit regression; the results of the probit regression are very similar and are available upon request. Here, again, we used the model with diffused information transmission for the same parameter values as in previous simulations (Sections 4 and 5.1), listed in the Appendix; the results for the focused transmission are qualitatively the same.

[14] Hashimzade et al. (2014) demonstrate that targeted audits also outperform mixed audit strategies for both groups in terms of the tax yield and average compliance but not necessarily in terms of compliance. The "mixed" regimes of targeted and random audits are akin to the random enquiry programmes run alongside targeted audits by the US IRS and the UK HMRC.

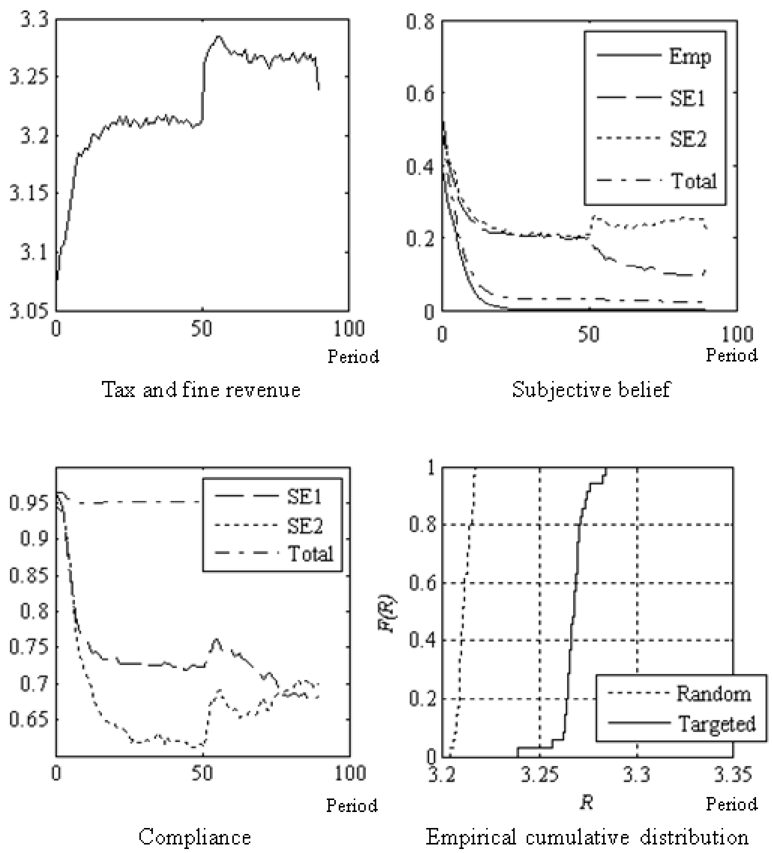

Figure 2: Subjective beliefs and compliance of all taxpayers, taxpayers in employment, in SE1 and in SE2; tax and fine revenue levels and distribution, under the diffused information transmission. The tax authority uses random audits up to period 50 and predictive analytics to target the largest evaders from period 51 onwards.

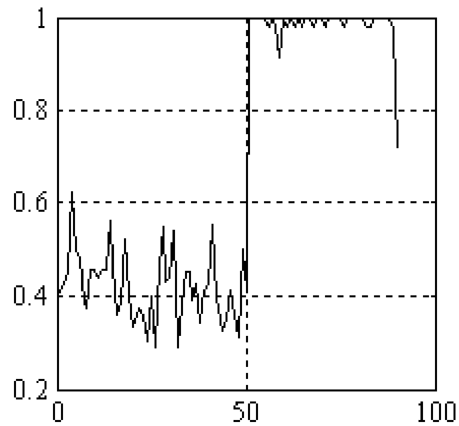

Figure 3: Proportion of SE2 among audited taxpayers in each period under the diffused information transmission. The tax authority implements random audits up to period 50 and predictive analytics to target the largest evaders from period 51 onwards.

The simulation results when the most likely evaders are targeted are shown in Figure 4. As in Figure 2, the values for each time period correspond to the averages over ten independent simulations.[15] By comparing Figure 4 with Figure 2 it may be seen that the evolution of revenue and beliefs when the most likely evaders are targeted is very similar to that when the largest evader are targeted. Compliance behavior, however, appears to become somewhat more sensitive to the around the introduction of predictive analytics, with a bigger fall in compliance from taxpayers in SE1 and a bigger rise in compliance from taxpayers in SE2. The empirical cdf for revenue when targeting the largest evaders and when targeting the most likely evaders are very close, which suggests that these strategies can be equally successful in recovering unpaid tax.

Table 4 shows the marginal effects of explanatory variables on the predicted probability of evasion. As in Section 5.1, the estimated coefficients change as new data are added. We report the estimates based on the data accumulated up to the final period in the last independent simulation; this allows capturing the long-run outcome and using meaningful standard errors. The table shows that individuals with higher declared income, those audited in the previous period, and those in SE2 are less likely to evade.[16] These results are similar to the predictions of the model for level-targeting; neither the marginal effect of the previous audit nor the effect of occupation are statistically significant at conventional levels.

[15] Across 90 periods, the ratios of the standard deviations to the means over five simulations ranged from 0.2% to 0.8% for revenues, 0.2% to 22.7% for compliance, and 0.4% to 69.4% for beliefs.

[16] For the binary variables the adjusted marginal effects are reported. That is, for example, the probability of an "average" SE1 taxpayer evading tax is by 6.3% lower than that of an "average" SE2 taxpayer.

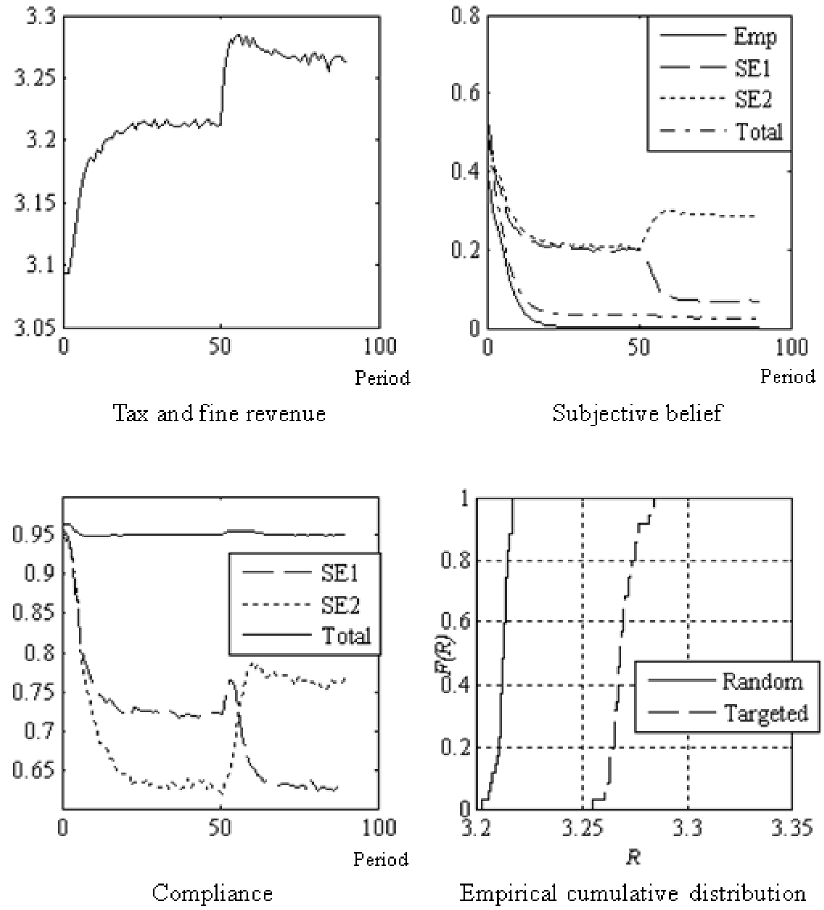

Figure 4: Subjective beliefs and compliance of all taxpayers, taxpayers in employment, in SE1 and in SE2; tax and fine revenue levels and distribution., under the diffused information transmission. The tax authority uses random audits up to period 50 and predictive analytics to target the most likely evaders from period 51 onwards.

Variable	Coeff.	St. E.	$z-$stat.	ME (avg. data)	ME (indiv. avg.)
Const.	5.63	0.38	14.8	–	–
Declared Income	-0.75	0.0.7	-11.1	-0.008	-0.001
Previous audit	-11.7	263	-0.04	-0.998	-0.922
SE1	-2.12	0.63	-3.33	-0.063	-0.031

Table 4: Estimated marginal effects from the evasion score equation (logit model). The reported coefficients and standard errors are the maximum likelihood estimates from the data accumulated up to the final period of one simulation (1943 observations).

6 Conclusions

The optimal design of audit strategy is important for tax authorities, whose aim is to design policy instruments to reduce the tax gap (the difference between anticipated and actual tax revenue). In this paper we analyze two alternative strategies that use the concept of predictive analytics: targeting the largest evaders and targeting the most likely evaders. We do this is in a rich network model where taxpayers are heterogeneous in risk, beliefs, and attitude towards compliance, and where agents self-select into different occupational groups. In this model attitudes and beliefs endogenously emerge that differ across sub-groups of the population, compliance behavior is different across occupational groups, and these effects are reinforced by the development of group-specific attitudes and beliefs. Given this behavior, the tax authority may wish to condition its audit strategy not only on reported income, but also on occupation.

What does our model suggest for the optimal strategy of a tax authority? On the one hand, given the objective of maximizing revenues, targeting the level, or the value, of evasion appears to be more important. On the other hand, the "strike rate", or the proportion of audits that reveal evasion, is also important, if the tax authority wants to reduce the burden on the compliant population (James, 2011). Our results imply that the two strategies have the same quantitative effect on the revenues; furthermore, the rates of compliance in population and in each occupation are not noticeably different between the two strategies.

The robustness of this conclusion to the selection of model assumptions is, however, an issue that could be addressed in future research. We have assumed that the agents live indefinitely long throughout the model. It would be possible to incorporate finite lives within the simulation with some agents leaving the economy each period and new agents joining. What would be harder to incorporate is forward-looking agents because solving the fixed-point problem that would then arise would be exceptionally demanding on computational resources.[17] We have adopted a very simple process of random network formation. Other network formation processes would be interesting, particularly networks with "celebrities" who influence many other agents. Another audit strategy of interest to investigate is the "light-touch" audit strategies, where either random or targeted audit can reveal only a fraction of concealed income but at a lower cost to the tax authority. The light-touch audits allow a wider coverage of population, thereby raising subjective beliefs and improving compliance; however, partial detection increases expected payoff from evasion and encourages non-compliance (Rablen, 2014). This trade-off, along with the cost considerations, can lead to the selection of an optimal mix of audits.

[17] We run the simulation on a computer with Intel multi-core and 64-bits. The computing time for ten indepedent simulations of the model is approximately 48 hours. Incorporating forward-looking agents would require a considerable reduction in population size in order to make the analysis computationally feasible.

References

Advani, A., Elming, W., Shaw, J., 2015. How long lasting are the effects of audits?, TARC Discussion Paper, 011-15.

Allingham, M., Sandmo, A., 1972. Income tax evasion: A theoretical analysis. J. Pub. Econ. 1, 323–338.

Alm, J., Erard, B., 2013. Using public information to estimate informal supplier income, unpublished, Tulane University.

al-Nowaihi, A., Dhami, S., 2007. Why do people pay taxes: Expected utility theory versus prospect theory. J. Econ. Behav. Org. 64, 171–192.

Andrei, A.L., Comer, K., Koelher, M., 2014. An agent-based model of network effects on tax compliance and evasion. J. Econ. Psychol. 40, 119–133.

Arcand, J.-L., Graziosi, G.R., 2005. Tax compliance and rank dependent expected utility. Geneva Risk Insur. Rev. 30, 57–69.

Ashby, J.S., Webley, P., Haslam, A.S., 2009. The role of occupational taxpaying cultures in taxpaying behaviour and attitudes. J. Econ. Psychol. 30, 216–227.

Axtell, R.L., Epstein, J.M., Young, H.P., 2006. The emergence of classes in a multi-agent bargaining model. In: Epstein, J.M. (Ed.), Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton, NJ.

Baldry, J.C., 1986. Tax evasion is not a gamble. Econ. Lett. 22, 333–335.

Bernasconi, M., 1998. Tax evasion and orders of risk aversion. J. Pub. Econ. 67, 123–134.

Bernasconi, M., Zanardi, A., 2004. Tax evasion, tax rates and reference dependence. FinanzArchiv 60, 422–445.

Black, J., de Meza, D., 1997. Everyone may benefit from subsidising entry to risky occupations. J. Pub. Econ. 66, 409–424.

Bloomquist, K.M., 2004. Multi-agent based simulations of the deterrent effects of taxpayer audits. In: Kalambokidis, L. (Ed.), Proceedings of the 97th Annual Conference on Taxation. National Tax Association, Washington, DC.

Bloomquist, K.M., 2011. Tax compliance as an evolutionary coordination game: An agent-based approach. Public Finan. Rev. 39, 25–49.

Bloomquist, K.M., 2013. Incorporating indirect effects in audit case selection: An agent-based approach. IRS Res. Bull. 103–116.

Bradbury, B., 1997. The living standards of the low income self-employed. Australian Econ. Rev. 30, 374–389.

Camerer, C., Weber, M., 1992. Recent development in modelling preferences: Uncertainty and ambiguity. J. Risk Uncertain. 5, 325–370.

Chander, P., Wilde, L.L., 1998. A general characterization of optimal income tax enforcement. Rev. Econ. Stud. 65, 165–183.

Chetty, R., 2009. Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. Amer. Econ. J.: Econ. Pol. 1, 31–52.

Christiansen, V., 1980. Two comments on tax evasion. J. Pub. Econ. 13, 389–393.

Cleary, D., 2011. Predictive analytics in the public sector: Using data mining to assist better target selection for audit. Electron. J. e-Govern. 9, 132–140.

Clotfelter, C.T., 1983. Tax evasion and tax rates: An analysis of individual returns. Rev. Econ. Stat. 65, 363–373.

Cowell, F.A., 1981. Taxation and labour supply with risky activities. Economica 48, 365–379.

Cowell, F.A., 1985. Tax evasion with labour income. J. Pub. Econ. 26, 19–34.

Cowell, F.A., Gordon, J.P.F., 1988. Unwillingness to pay. J. Pub. Econ. 36, 305–321.

Crane, S.E., Nourzad, F., 1986. Inflation and tax evasion: An empirical analysis. Rev. Econ. Stat. 68, 217–223.

Davis, J.S., Hecht, G., Perkins, J.D., 2003. Social behaviours, enforcement, and tax compliance dynamics. Account. Rev. 78, 39–69.

De Juan, A., Lasheras, M.A., Mayo, R., 1994. Voluntary tax compliant behavior of Spanish income tax payers. Public Finan. 49, 90–115.

Eide, E., 2001. Rank Dependent Expected Utility Models of Tax Evasion. Working Paper ICER 27/2001, International Centre for Economic Research.

Eisenhauer, J.G., 2006. Conscience as a deterrent to free riding. Int. J. Social Econ. 33, 534–546.

Eisenhauer, J.G., 2008. Ethical preferences, risk aversion, and taxpayer behavior. J. Socio. Econ. 37, 45–63.

Elffers, H., Hessing, D.J., 1997. Influencing the prospects of tax evasion. J. Econ. Psychol. 18, 289–304.

Feldman, N.E., Slemrod, J., 2007. Estimating tax noncompliance with evidence from unaudited tax returns. Econ. J. 117, 327–352.

Fortin, B., Lacroix, G., Villeval, M.-C., 2007. Tax evasion and social interactions. J. Pub. Econ. 91, 2089–2112.

Franklin, J., 2009. Networks and taxpayer non-compliance. London: H.M. Revenue and Customs.

Geeroms, H., Wilmots, H., 1985. An empirical model of tax evasion and tax avoidance. Public Finan. 40, 190–209.

Glaeser, E., Sacerdote, B., Scheinkman, J., 1996. Crime and social interactions" Q. J. Econ. 111, 507–548.

Gordon, J.P.F., 1989. Individual morality and reputation costs as deterrents to tax evasion. Eur. Econ. Rev. 33, 797–805.

Goyal, S., 2009. Connections: An Introduction to the Economics of Networks. Princeton University Press, Princeton, NJ.

Grether, D., 1980. Bayes rules as a descriptive model: The representativeness heuristic. Q. J. Econ. 95, 537–557.

Guala, F., Mittone, L., 2005. Experiments in economics: External validity and the robustness of phenomena. J. Econ. Method. 12, 495–515.

Hamilton, B.H., 2000. Does entrepreneurship pay? An empirical analysis of the returns to self employment, J. Pol. Econ., 108, 604–631.

Hashimzade, N., Myles, G.D., Rablen, M.D., 2014. Predictive analytics and the targeting of audits, TARC Discussion Paper, 008-14.

Hashimzade, N., Myles, G.D., Tran-Nam, B., 2013. Applications of behavioural economics to tax evasion, J. Econ. Surv. 27, 941–977.

Hsu, K.-W., Pathak, N., Srivastava, J., Tschida, G., Bjorklund, E., 2015. Data mining based tax audit selection: A case study of a pilot project at the Minnesota Department of Revenue. Ann. Inform. Syst. 17, 221–245.

IRS, 2014. Internal Revenue Service Data Book, 2013. Publication 55B. IRS, Washington, DC.

Isachsen, A.J., Strøm, S., 1980. The hidden economy: The labor market and tax evasion. Scand. J. Econ. 82, 304–311.

Jackson, M.O., 2004. A survey of models of network formation: Stability and efficiency. In: Demange, G., Wooders, M. (Eds.), Group Formation in Economics; Networks, Clubs and Coalitions. Cambridge University Press: Cambridge.

James, R., 2011. Using predictive analytics to help target error and fraud in tax credits. International Conference on Taxation Analysis and Research, London, December 2011.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–293.

Kanbur, S.M., 1981. Risk taking and taxation: An alternative perspective. J. Pub. Econ. 15, 163–184.

Kastlunger, B., Kirchler, E., Mittone, L., Pitters, J., 2009. Sequences of audits, tax compliance, and taxpaying strategies. J. Econ. Psychol. 30, 405–418.

Kim, Y., 2003. Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. J. Pub. Econ. 87, 1591–1616.

Kirchler, E., Hoelzl, E., Wahl, I., 2008. Enforced versus voluntary tax compliance: The "slippery slope" framework. J. Econ. Psychol. 29, 210–225.

Kolm, S.-C., 1973. A note on optimum tax evasion. J. Pub. Econ. 2, 265–270.

Korobow, A., Johnson, C., Axtell, R., 2007. An agent-based model of tax compliance with social networks. Natl. Tax J. 60, 589–610.

Lee, K., 2001. Tax evasion and self-insurance. J. Pub. Econ. 81, 73–81.

Maciejovsky, B., Kirchler, E., Schwarzenberger, H., 2007. Misperception of chance and loss repair: On the dynamics of tax compliance. J. Econ. Psychol. 28, 678–691.

Mittone, L., 2006. Dynamic behaviour in tax evasion: An experimental approach. J. Socio. Econ. 35, 813–835.

Myles, G.D., 1995. Public Economics. Cambridge University Press, Cambridge.

Myles, G.D., Naylor, R. A., 1996. A model of tax evasion with group conformity and social customs. Eur. J. Polit. Econ. 12, 49–66.

Pestieau, P., Possen, U.M., 1991. Tax evasion and occupational choice. J. Pub. Econ. 45, 107–125.

Pissarides, C.A., Weber, G., 1989. An expenditure-based estimate of Britain's black economy. J. Pub. Econ. 39, 17–32.

Plumley, A.H., Steuerle, C.E., 2004. Ultimate objectives for the IRS: Balancing revenue and service. In: Aaron, H.J., Slemrod, J. (Eds.), The Crisis in Tax Administration. Brookings Institution Press, Washington DC.

Pyle, D.J., 1991. The economics of taxpayer compliance. J. Econ. Surv. 5, 163–198.

Quiggin, J., 1981. Risk perception and the analysis of risk attitudes. Australian J. Agric. Econ. 25, 160–169.

Quiggin, J., 1982. A theory of anticipated utility. J. Econ. Behav. Organ. 3, 323–343.

Quiggin, J., Wakker, P., 1994. The axiomatic basis of anticipated utility: A clarification. J. Econ. Theory 64, 486–499.

Rablen, M.D., 2010. Tax evasion and exchange equity: A reference-dependent approach. Public Finan. Rev. 38, 282–305.

Rablen, M.D., 2014. Audit probability versus effectiveness: The Beckerian approach revisited. J. Public Econ. Theor. 16, 322–342.

Reinganum, J., Wilde, L., 1985. Income tax compliance in a principal-agent framework. J. Pub. Econ. 26, 1–18.

Reinganum, J., Wilde, L., 1986. Equilibrium verification and reporting policies in a model of tax compliance. Int. Econ. Rev. 27, 739–760.

Sandmo, A., 2005. The theory of tax evasion: A retrospective view. Natl. Tax J. 58, 643–663.

Schneider, F., Enste, D.H., 2000. Shadow economies: Size, causes, and consequences. J. Econ. Lit. 38, 77–114.

Slemrod, J., 2007. Cheating ourselves: The economics of tax evasion. J. Econ. Perspect. 21, 25–48.

Snow, A., Warren, R.S. Jr., 2005. Ambiguity about audit probability, tax compliance, and taxpayer welfare. Econ. Inq. 43, 865–871.

Souma, W., Fujiwara, Y., Aoyama, H., 2003. Complex networks and economics. Physica A 324, 396–401.

Spicer, M.W., Lundstedt, S.B., 1976. Understanding tax evasion. Public Finan. 31, 295–305.

Tesfatsion, L., 2006. Agent-based computational economics: A constructive approach to economic theory. In: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics. North-Holland, Amsterdam.

Thaler, R.H., 1994. The winner's curse: Paradoxes and anomalies of economic life. Princeton University Press, Princeton, NJ.

Torgler, B., 2002. Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. J. Econ. Surv. 16, 657–683.

Trandel, G., Snow, A., 1999. Progressive income taxation and the underground economy. Econ. Lett. 62, 217–222.

Traxler, C., 2010. Social norms and conditional cooperative taxpayers. Eur. J. Polit. Econ. 26, 89–103.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323.

Wallschutzky, I.G., 1984. Possible causes of tax evasion. J. Econ. Psychol. 5, 371–384.

Webley, P., Robben, H., Morris, I., 1988. Social comparison, attitudes and tax evasion in a shop simulation. Social Behav. 3, 219–228.

Weigel, R.H., Hessing, D.J., Elffers, H., 1987. Tax evasion research: A critical appraisal and theoretical model. J. Econ. Psychol. 8, 215–235.

Yaniv, G., 1999. Tax compliance and advance tax payments: A prospect theory analysis. Natl. Tax J. 52, 753–764.

Yitzhaki, S., 1974. A note on income tax evasion: A theoretical analysis. J. Pub. Econ. 3, 201–202.

Parameter values
Probability of being linked in the network: $v = 0.5$;
Tax rate: $\tau = 0.25$;
Fine rate: $f = 1.5$;
Weight in information exchange: $\mu = 0.75$;
Probability of meeting a network contact: $\phi = 0.25$;
Probability of information exchange
  Focussed information transmission: $q^{\alpha\beta} = 0$, $q^{\alpha\alpha} = 1$ for $\alpha \neq \beta$, $\alpha, \beta = \{0, 1, 2\}$;
  Diffused information transmission: $q^{\alpha\beta} = 0.15$, $q^{\alpha\alpha} = 0.75$ for $\alpha \neq \beta$, $\alpha, \beta = \{0, 1, 2\}$;
Number of agents: $N = 6000$.
Time horizon: $T = 90$.

Probability distributions ($\mathcal{N}$ = normal; $\mathcal{U}$ = continuous uniform)
Wage in employment: $\log w \sim \mathcal{N}(\mu_0, \sigma_0^2)$;
  $E[w] = 13.0425$; $Var(w) = 4$;
Skill in SE1:
  $s_j^1 = \frac{1}{1-0.5\tilde{x}}$, $\tilde{x} = \mathcal{U}[0, 1]$;
  $E\left[s_j^1\right] = 1.387$; $Var\left(s_j^1\right) = 0.078$;
Income in SE1:
  $s_j^1 y^1$, $\log y^1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$;
  $E[y^1] = 8.01$; $Var(y^1) = 12.26$;
Skill in SE2:
  $s_j^2 = \frac{1}{1-0.5\tilde{x}}$, $\tilde{x} = \mathcal{U}[0, 1]$;
  $E\left[s_j^2\right] = 1.387$; $Var\left(s_j^2\right) = 0.078$;
Income in SE2:
  $s_j^2 y^2$, $\log y^2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$;
  $E[y^2] = 8.3$; $Var(y^1) = 22.8$;
Risk aversion: $\rho \sim \mathcal{U}[0.1, 5.1]$;
Initial belief on audit probability: $p_0 \sim \mathcal{U}[0, 1]$;
Initial weight on the payoff to following the social custom: $\chi_0 \sim \mathcal{U}[0, 1]$;
Value of social custom: $z \sim 3 \times 10^{-5} \times \mathcal{U}[0, 1]$.